ANNALS OF PHYSICS **191**, 363–381 (1989)

Optimal State-Determination by Mutually Unbiased Measurements

WILLIAM K. WOOTTERS AND BRIAN D. FIELDS

Department of Physics, Williams College, Williamstown, Massachusetts 01267

Received October 26, 1988

For quantum systems having a finite number N of orthogonal states, we investigate a particular relation among different measurements, called "mutual unbiasedness," which we show plays a special role in the problem of state determination. We define two bases $\{|v_i\rangle\}$ and $\{|w_j\rangle\}$ to be mutually unbiased if all inner products across their elements have the same magnitude: $|\langle v_i|w_j\rangle| = 1/\sqrt{N}$ for all i, j. Two non-degenerate measurements are defined to be mutually unbiased if the bases comprising their eigenstates are mutually unbiased. We show that if one can find $N+1$ mutually unbiased bases for a complex vector space of N dimensions, then the measurements corresponding to these bases provide an optimal means of determining the density matrix of an ensemble of systems having N orthogonal states, in the sense that the effects of statistical error are minimized. We show further that the number of mutually unbiased bases one may find for a given N is at most $N+1$. Finally, we show that $N+1$ mutually unbiased bases do exist whenever N is a power of a prime, and we construct such bases explicitly.

INTRODUCTION

Quantum mechanics adopts the formalism of the complex vector space as a convenient means of expressing quantum states and operators. Thus one often finds a close correspondence between interesting features of this mathematical framework and important physical properties. For example, one crucial feature of the mathematics is that vectors can have the property of orthogonality, and this property appears naturally in the quantum formalism: the eigenstates of a non-degenerate measurement form an orthogonal basis for the space they occupy. Other bases are mathematically possible, but the formalism of quantum mechanics favors the special bases which are orthogonal. In this paper we examine a particular relation among *different* orthogonal bases, a relation we call "mutual unbiasedness." We intend to show that just as orthogonal bases are a special and important type of basis for describing states, in a somewhat analogous way "mutually unbiased" bases are a favored case of orthogonal bases. We consider particularly the special role that mutually unbiased bases play in the problem of determining the state of a quantum ensemble. As we shall see, measurements associated with mutually unbiased bases provide an optimal means of determining an ensemble's

0003-4916/89 $7.50

state. In this paper we restrict our attention to vector spaces of finite dimension, so our work applies to quantum systems having only a finite number of orthogonal states.

We define the bases $\{|v_i\rangle\}$ and $\{|u_j\rangle\}$ over an N-dimensional complex space to be mutually unbiased if inner products between all possible pairs of vectors with one vector from each basis all have the same magnitude:

$$|\langle v_i|u_j\rangle| = \frac{1}{\sqrt{N}}. \tag{1}$$

One can alternatively express the notion of mutual unbiasedness in terms of projection operators. Let P_i be the projection onto the vector $|v_i\rangle$ and let Q_j project onto $|u_j\rangle$. Then mutual unbiasedness requires that for all P_i and Q_j,

$$\mathrm{Tr}(P_i Q_j) = \frac{1}{N}. \tag{2}$$

The example of such bases which is simplest and most widely known is for $N=2$. Three $N=2$ bases that are easily verified to be mutually unbiased are

$$\left\{\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right\}, \quad \left\{\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}\right\}, \quad \left\{\frac{1}{\sqrt{2}}\begin{pmatrix}1\\i\end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-i\end{pmatrix}\right\}. \tag{3}$$

For a spin-$\frac{1}{2}$ particle, these bases are, in a standard representation, the eigenstates of the components of spin in the z, x, and y directions, respectively.

We will use the expression "mutually unbiased" to apply to measurements as well as to bases: two non-degenerate measurements are mutually unbiased if the bases comprising their eigenstates are mutually unbiased. Thus the measurements of components of spin along the x, y, and z axes are all unbiased with respect to each other.

Pairs of bases related in this way for finite N were discussed by Schwinger, who noted that these bases represent measurements that are "maximally noncommutative" [1]. In other words, a measurement over one basis leaves one completely uncertain as to the outcome of a measurement over a basis unbiased with respect to the first. For example, a measurement of spin in the x direction for a spin-$\frac{1}{2}$ particle leaves one completely uncertain as to the component of spin in the y direction. In part because of this maximal incompatibility, Schwinger called operators corresponding to such bases "complementary." Recently, Kraus also examined such bases and speculated that they correspond to measurements which yield the strongest conceivable bound on an information-theoretic statement of the uncertainty relation [2]. Massen and Uffink have since verified Kraus' conjecture [3].

Whereas others have focused attention on pairs of unbiased bases, Ivanovic has examined sets of several bases all of which are pairwise mutually unbiased [4]. He discussed such bases in the context of the problem of state determination, which is also the main focus of the present paper.

The problem is this: to ascertain the density matrix of an ensemble of identical N-state systems, which may be in a pure or a mixed state. Such a density matrix is specified by N^2-1 real parameters [5]. Any given non-degenerate measurement applied to a subensemble will yield precisely $N-1$ real numbers, namely, the probabilities of all but one of the N possible outcomes, the last probability being determined by the requirement that the probabilities sum to unity. Thus the minimum number of different measurements one needs in order to determine the state uniquely is $(N^2-1)/(N-1)$, that is, $N+1$. Ivanovic has shown that if one can find $N+1$ bases that are all unbiased with respect to each other, then the measurements corresponding to these bases are guaranteed to be sufficient for determining the density matrix.

For the $N=2$ case, the sufficiency of $N+1$ mutually unbiased bases for reconstructing the density matrix is clear. The density matrix in that case can be written in the form $W=(I+\mathbf{r}\cdot\boldsymbol{\sigma})/2$, where I is the unit matrix and $|\mathbf{r}|\leqslant 1$. The measurements represented by the Pauli matrices σ_x, σ_y, and σ_z are mutually unbiased, and each of these measurements, applied to a large subensemble, determines precisely one component of the vector $\mathbf{r}$. Thus if the three measurements are applied to three distinct subsets of the original ensemble, one obtains all three components of $\mathbf{r}$, and hence the density matrix W.

It is clear from this example that one can also determine the state by means of $N+1$ measurements that are *not* mutually unbiased. For an ensemble of spin-$\frac{1}{2}$ particles, it is sufficient to use measurements of spin along any three non-coplanar axes: each measurement, performed on a subensemble, restricts the vector $\mathbf{r}$ to a specific plane, and the intersection of the three planes determines $\mathbf{r}$ uniquely. However, there is a sense in which the unbiased case is the best. If the ensemble is finite, then one will not be able to avoid statistical error; that is, one will not be able to obtain with absolute precision the probabilities of the outcomes of any given measurement. In the spin example this means that each measurement confines $\mathbf{r}$ not to a plane but to a "fuzzy plane," which can be pictured as a pancake-shaped region. The intersection of three such pancakes is not a point but has some non-zero volume. As long as the widths of the three pancakes are fixed, it is clear that the volume of the intersection, that is, the final uncertainty in the vector $\mathbf{r}$, is minimized if the three pancakes are mutually perpendicular, that is, if the three measurements are mutually unbiased.

In fact the situation is more complicated than this: in the case of statistical error, the widths of the pancakes are not independent of the choice of measurement. In the following section we address this problem, and we prove that in a precisely defined sense, a set of $N+1$ mutually unbiased measurements provides the optimal determination of an unknown state, provided that such measurements exist.

The remainder of the paper addresses the question of the existence of $N+1$ mutually unbiased bases. Ivanovic has shown by explicit construction that such bases exist for each prime value of N. We extend this result to all values of N which are powers of primes. We also show that for no value of N is it possible to find *more* than $N+1$ mutually unbiased bases.

Optimal State Determination

We now consider the problem of state determination for an arbitrary value of N. Let us begin by stating the problem carefully. We are given a large but finite ensemble of identical systems, each having N orthogonal states, whose density matrix is to be determined. The set of all possible density matrix is the set of all trace-one Hermitian matrices with non-negative eigenvalues. For our purposes it will be more convenient to represent states by *traceless* Hermitian matrices, the state with density matrix W being represented by $Y_W = W - I/N$ rather than by W itself. Let S be the set of all such Y's. S is a subset of the set of *all* traceless Hermitian matrices, which we call T. The set T has the convenient property of being a vector space under ordinary matrix addition, and this is the main reason we are using traceless matrices to represent states. Moreover, the vector space T comes equipped with a natural inner product, given by the trace of the matrix product. That is, for any two matrixes A and B in T, the inner product between A and B is $\mathrm{tr}(AB)$. The dimension of T is $N^2 - 1$. Later we will be doing integrals over S, and the notion of volume we will use, i.e., the measure, is the one induced by the inner product: the inner product defines length and angle, and from these one can compute volume. For example, the volume of a hypercube whose edges have unit length is one.

Our problem can now be restated as follows. At the beginning of the experiment, the ensemble's state could be anywhere in the set S. More precisely, we assume that the a priori distribution is the uniform distribution relative to the natural measure on the vector space T. We now perform a series of non-degenerate measurements on the ensemble. Specifically, we divide the ensemble into $N+1$ subensembles of equal size, and on each of these subensembles we perform a different measurement. From the probabilities of the outcomes of these measurements we attempt to reconstruct the original state of the ensemble. However, because of statistical error we can know these probabilities only approximately, so the amount of information we can gain about the ensemble's original state is limited. Our object is to choose the $N+1$ measurements in such a way as to maximize the information we gain.

How does one quantify this information? The information from the various measurements adds to one's knowledge of the location of the actual state within the entire state space S. This knowledge is expressed as a probability distribution over the state space, each measurement contributing to the final distribution. (Throughout this section, when we speak of a "measurement," we mean a measurement performed on an entire subensemble.) As we shall see, this final probability distribution of states turns out to be a Gaussian distribution over the set S. Such a distribution can be characterized by an ellipsoidal region E defined as follows: E is the set of all states for which the probability density exceeds $1/e$ times its maximum value. One could use the volume of E to characterize the overall spread of the distribution, but the equations will be simpler if we use instead the volume V of the smallest rectangular parallelepiped that encloses E. We will refer to V as the "uncertainty volume" of the distribution. According to the usual definition

of information, as applied to a continuous distribution, the amount of information gained in going from the uniform a priori distribution to the final Gaussian distribution is given by [6]

$$\mathscr{I} = -\ln\left(\frac{V}{V_0}\right) - \left(\frac{N^2-1}{2}\right)\ln(\pi e), \tag{4}$$

where V_0 is the volume of the set S. It will turn out that this information depends on the state of the ensemble. That is, the same measurement scheme can yield more or less information, depending on the ensemble's state. Let $\mathscr{I}(Y)$ be the amount of information one gains when the actual state is Y (Y is an element of S; the corresponding density matrix is $Y + I/N$). The average information $\langle \mathscr{I} \rangle$ one may expect to gain in the course of the measurements is given by $(1/V_0)\int \mathscr{I}(Y)\, dY$, where dY indicates integration with respect to the natural measure on S. It is this average information that we wish to maximize.

We now imagine performing one of the $N+1$ measurements. Let $\{P_i^{(r)}\}$, $i = 1, ..., N$, be the operators that project onto the eigenstates of the rth measurement. When we perform this measurement on the rth subensemble, the probabilities of the N outcomes are $p_i = \mathrm{tr}(WP_i^{(r)})$, where W is the ensemble's density matrix. Suppose for a moment that we could determine these p_i's exactly. How would that information restrict the set of possible density matrices? It is easier to work with the numbers $q_i = p_i - 1/N$. They have the property that

$$q_i = \mathrm{tr}(YP_i^{(r)}) = \mathrm{tr}[\,Y(P_i^{(r)} - I/N)], \tag{5}$$

where $Y = W - I/N$ is the matrix that represents the ensemble's state in S. The matrices $P_i^{(r)} - I/N$, whose sum is zero, span an $(N-1)$-dimensional subspace of T, which we call T_r. According to Eq. (5), the q_i's determine the projection of Y on the subspace T_r. Thus if the p_i's, and hence the q_i's, were known exactly, the ensemble's state would be restricted to the set of all states having this projection. This set is a flat space of $N^2 - N$ dimensions which is perpendicular to T_r and which intersects T_r at a point determined by the values of the q_i's.

In actuality we cannot determine the q_i's exactly. Because the ensemble is finite, the observed frequencies of occurrence of the N outcomes will not be exactly equal to the probabilities of these outcomes. Rather, they will be distributed around these probabilities according to a multinomial distribution. For a sufficiently large ensemble, such a distribution can be well approximated by a Gaussian [7]. Thus our knowledge of the q_i's will be given by a Gaussian distribution over the $(N-1)$-dimensional space defined by $\sum q_i = 0$ [8]. This implies that the projection of Y on the subspace T_r will not be restricted to a point, but rather will be spread according to a Gaussian distribution. The uncertainty volume in T_r associated with this distribution, defined as above, is given by $V_r = (2/m)^{[(N-1)/2]}\sqrt{N}\sqrt{p_1 \cdots p_N}$, where m is the number of elements of the subensemble being measured [7]. Thus we can picture the result of the rth measurement as follows: our knowledge of Y is

characterized by a probability distribution which is constant along the directions perpendicular to the subspace T_r, and which within T_r is a Gaussian having an overall spread V_r.

The measurement of each subensemble, taken by itself, will have such an effect. That is, it restricts the state according to a distribution that is narrow in $N-1$ dimensions and essentially infinitely broad in the other N^2-N dimensions. When the results of all the measurements are taken together, the final probability distribution for the ensemble's state is the product of the distributions determined by the $N+1$ individual measurements. This final distribution wil be a Gaussian, and as long as the measurements have been chosen so that the subspaces T_r are linearly independent, there will no longer be any dimensions in which the state is unrestricted.

One can show that the overall spread of this final distribution is related to the spreads of the individual distributions according to the equation

$$V=\frac{V_1\cdots V_{N+1}}{\mathrm{vol}(T_1,\ldots,T_{N+1})}. \tag{6}$$

Here $\mathrm{vol}(T_1,\ldots,T_{N+1})$ is the volume of an (N^2-1)-dimensional parallelepiped constructed as follows: choose any orthonormal basis for each of the spaces $T_1,\ldots,T_{N+1}$, and let the elements of these bases be edges of the parallelepiped. The volume of this parallelepiped depends only on the geometrical relationships among the subspaces.

According to Eqs. (4) and (6), the information we have gained by performing the $N+1$ measurements is

$$\mathscr{I}=-\sum_{r=1}^{N+1}\ln(V_r)+\ln[\mathrm{vol}(T_1,\ldots,T_{N+1})]+\ln(V_0)-\left(\frac{N^2-1}{2}\right)\ln(\pi e). \tag{7}$$

In this expression, only the first term depends on the state of the ensemble. The second term depends only on the choice of measurements, and the last two terms are constants. Thus, when we average over all possible initial states, we get

$$\langle\mathscr{I}\rangle=-\sum_{r=1}^{N+1}\langle\ln(V_r)\rangle+\ln[\mathrm{vol}(T_1,\ldots,T_{N+1})]+\ln(V_0)-\left(\frac{N^2-1}{2}\right)\ln(\pi e), \tag{8}$$

where, as before, $\langle\cdots\rangle$ indicates an average over the set S. Now the average $\langle\ln(V_r)\rangle$ cannot depend on the actual eigenvectors of the rth measurement, for the following reason. The set of states over which one is averaging, as well as the measure used in taking the average, is invariant under unitary transformations. The bases corresponding to two different measurements are always related by a unitary transformation (neglecting the ordering of the basis elements), so the average $\langle\ln(V_r)\rangle$ must be the same for all measurements. This means that maximizing the

information boils down to maximizing the quantity $\mathrm{vol}(T_1, ..., T_{N+1})$, this being the only term in Eq. (8) that depends on how we choose our measurements.

We thus wish to maximize $\mathrm{vol}(T_1, ..., T_{N+1})$, the volume of a parallelepiped whose edges are all of unit length. If one were allowed to choose the subspaces $T_1, ..., T_{N+1}$ freely, without forcing them to be related to measurements, then this volume would be maximized if the unit vectors defining the parallelepiped were all perpendicular, and this would happen only if the subspaces $T_1, ..., T_{N+1}$ were mutually orthogonal. We now show that these subspaces will indeed be orthogonal if the $N+1$ measurements that define them are mutually unbiased.

The subspace T_r corresponding to the rth measurement is the space spanned by the operators $P_i^{(r)} - I/N$. We wish to show that if the measurements are mutually unbiased, then any vector in T_r is orthogonal to any vector in T_s as long as r is different from s. It is sufficient to show that $P_i^{(r)} - I/N$ is orthogonal to $P_j^{(s)} - I/N$ for all $r \neq s$, as we now do:

$$\mathrm{tr}[(P_i^{(r)} - I/N)(P_j^{(s)} - I/N)] = \mathrm{tr}(P_i^{(r)} P_j^{(s)}) - 2/N + 1/N = 0, \tag{9}$$

since $\mathrm{tr}(P_i^{(r)} P_j^{(s)}) = 1/N$ for unbiased bases.

Thus, if there exist $N+1$ mutually unbiased measurements, then the subspaces T_r associated with them are orthogonal, and therefore the measurements provide the maximum possible information about the ensemble's state. If there do not exist $N+1$ such measurements, then it will not be possible to make the subspaces $T_1, ..., T_{N+1}$ orthogonal to each other, and the problem of maximizing the information with $N+1$ measurements will be considerably more complicated. It is of interest, therefore, to investigate whether such measurements exist.

Note that Eq. (9) shows that there can never be *more* than $N+1$ mutually unbiased bases. If there were, then each one would still define an $(N-1)$-dimensional subspace T_r as above, and these subspaces would be orthogonal. But the number of dimensions of the whole space T is $N^2 - 1 = (N-1)(N+1)$, which is just large enough to accommodate $N+1$ such subspaces but no more. In fact, once one has $N+1$ mutually unbiased bases, it is not possible to find even a single vector that is unbiased with respect to all those bases: if such a vector $|v\rangle$ existed, then the operator $|v\rangle\langle v| - I/N$, regarded as a vector in T, would be orthogonal to all of the subspaces $T_1, ..., T_{N+1}$, and again this would contradict the fact that T has only $N^2 - 1$ dimensions. Because of this inextensibility of a set of $N+1$ mutually unbiased bases, we refer to such a set as "complete."

A Construction of Complete Sets of Mutually Unbiased Bases

We now construct complete sets of mutually unbiased bases for all values of N which are powers of primes. It turns out that the construction is different for powers of odd primes than for powers of two. Therefore we treat the two cases separately.

Unbiased Bases for Powers of Odd Primes

Ivanovic has shown that for $N=p$, an odd prime, one can find $N+1$ mutually unbiased bases [4]. These bases are the standard basis

$$(v_k^{(0)})_l = \delta_{kl}, \qquad k, l = 0, 1, ..., p-1 \tag{10}$$

(with the notation convention that the superscript denotes the basis, k the vector in the basis, l the component) and N bases in which the components of the vectors all have magnitude $1/\sqrt{N}$ and a phase that is a pth root of unity:

$$(v_k^{(r)})_l = \frac{1}{\sqrt{N}} e^{(2\pi i/p)(rl^2+kl)}, \qquad r = 1, 2, ..., N. \tag{11}$$

One can verify explicitly that each basis is orthonormal, and that the bases are mutually unbiased. In showing this last property, it is useful to employ the fact from number theory [4] that

$$\left| \sum_{j=0}^{p-1} e^{(2\pi i/p)(mj^2+nj)} \right| = \sqrt{p} \qquad (m \neq 0,\ p \text{ an odd prime}). \tag{12}$$

The key to generalizing this result is to note that the phases of the vectors are all pth roots of unity, so that the integral indices in the exponent have induced upon them an algebra modulo p. That p is prime gives this kind of algebra much more structure than an arbitrary arithmetic modulo h, where h is composite. Specifically, arithmetic modulo h is a ring in abstract algebra, but because of the indivisibility of primes, arithmetic modulo p has more structure than just a ring, and in fact is the simplest example of a field (using the word "field," as we do throughout, in its abstract algebraic sense) [9]. Recognizing the fundamental importance of the field algebra for the case $N=p$, one can generalize the above bases to the case of any power of an odd prime by letting the indices take values in fields with p^n elements, n a positive integer.

Written in field-theoretic notation, the generalization for $N=p^n$ $(p \neq 2)$ is as follows: As with the case of $N=p$, one can find $N+1$ bases, the first of which is again the standard one

$$(v_k^{(0)})_l = \delta_{kl}, \qquad k, l \in \mathbb{F}_{p^n}, \text{ the finite field with } p^n \text{ elements.} \tag{13}$$

The other N bases, which we will call the nonstandard bases, are given by

$$(v_k^{(r)})_l = \frac{1}{\sqrt{N}} e^{(2\pi i/p)\mathrm{Tr}(rl^2+kl)}, \qquad r, k, l \in \mathbb{F}_{p^n}, \quad r \neq 0. \tag{14}$$

Here Tr indicates the trace in the field theoretic sense, which will be defined in

Eq. (16). The demonstration that these are unbiased vectors relies on the result from field theory that [10]

$$\left| \sum_{j \in \mathbb{F}} e^{(2\pi i/p)\mathrm{Tr}(mj^2 + nj)} \right| = \sqrt{p^n} \qquad (m \neq 0,\ p \text{ an odd prime}). \tag{15}$$

One can use Eq. (15) to verify that the inner product between two of the above v's from different bases (i.e., different r's) has the right magnitude.

Even though Eqs. (13) and (14) provide a complete set of mutually unbiased bases for powers of odd primes and thus solve our problem for those values of N, it is worth examining these equations in more detail, partly in order to be able to write down the bases without using field elements explicitly, and partly in order to extend our result to the case $N = 2^n$. Clearly if one is to make sense of Eqs. (13) and (14), one needs to use the properties of finite fields, so we now summarize some basic definitions and theorems concerning these fields [11].

The simplest kind of field is that of the integers modulo p, p being any prime. This is called the Galois field $\mathbb{F}_p$ and contains the p elements $0, 1, \ldots, p-1$. From the Galois field one can construct fields having p^n elements, n a positive integer, as we show below. It is a theorem that the number of elements of a finite field *must* be a power of a prime, and that there exists only one field with any given number of elements p^n. The field with p^n elements, denoted $\mathbb{F}_{p^n}$, is called an extension of $\mathbb{F}_p$ and can be constructed by a process called root adjunction:

The strategy for constructing an extension field is similar to that for extending the real numbers to the complex numbers. In the latter case, one introduces a number i that solves the equation $i^2 = -1$, which is insoluble in the reals. The complex numbers are then the set of linear combinations of 1 and i, with real coefficients. One begins the analogous process for finite fields by finding an nth-degree polynomial with coefficients in $\mathbb{F}_p$ such that it cannot be factorized in $\mathbb{F}_p$ and as a consequence will have no roots in $\mathbb{F}_p$. One then introduces an element θ that is a root of the polynomial, and multiplicative closure requires that one also introduce the powers of θ: $\theta^2, \theta^3, \ldots, \theta^{n-1}$. One can then take linear combinations of the n powers of θ, $\{\theta^j\}$, and the set of all such combinations with coefficients in $\mathbb{F}_p$ forms a field with p^n elements. It turns out that other choices of the polynomial lead to the same structure, so that one may speak of *the* field with p^n elements. One can think of $\mathbb{F}_{p^n}$ as an n-dimensional vector space over $\mathbb{F}_p$, with the powers of θ forming a basis for the vector space. Any linearly independent combination of the powers of θ will serve equally well as a basis.

As an example of an extension field, consider the field for $p^n = 3^2$, recalling that $\mathbb{F}_3 = \{0, 1, 2\}$. A second-degree polynomial which is irreducible in $\mathbb{F}_3$ is $x^2 + x + 2 = 0$. Taking θ to be a root of this equation, we find that the nine elements of $\mathbb{F}_9$ are $\{0, 1, 2, \theta, \theta+1, \theta+2, 2\theta, 2\theta+1, 2\theta+2\}$. One can construct addition and multiplication tables for $\mathbb{F}_9$ using the fact that $\theta^2 + \theta + 2 = 0$.

The trace operator Tr that we used in Eq. (14) is defined as

$$\mathrm{Tr}\,\alpha = \alpha + \alpha^p + \alpha^{p^2} + \cdots + \alpha^{p^{n-1}}, \qquad \text{for all} \quad \alpha \in \mathbb{F}_{p^n}. \tag{16}$$

While the trace is in practice difficult to calculate, it has several important properties:

(1) $\mathrm{Tr}\,\alpha \in \mathbb{F}_p$ for all $\alpha \in \mathbb{F}_{p^n}$.

(2) $\mathrm{Tr}\,\alpha$ is linear in $\mathbb{F}_{p^n}$, where scalars are elements of $\mathbb{F}_p$.

(3) The linear transformations from $\mathbb{F}_{p^n}$ to $\mathbb{F}_p$ are exactly the mappings

$$\alpha \to \mathrm{Tr}(\beta\alpha), \qquad \beta \in \mathbb{F}_{p^n}.$$

With the field-theoretic tools now available, we may work to rewrite the bases for $N = p^n$ in a form not containing field elements explicitly.

In terms of field elements, the nonstandard bases are given by

$$(v_k^{(r)})_l = \frac{1}{\sqrt{N}} e^{(2\pi i/p)\mathrm{Tr}(rl^2+kl)}, \qquad r, k, l \in \mathbb{F}_{p^n}, \quad r \neq 0. \tag{17}$$

Consider the argument of the trace in the exponent, $rl^2 + kl$. Since we may think of $\mathbb{F}_{p^n}$ as an n-dimensional vector space, we may write the elements of $\mathbb{F}_{p^n}$ as linear combinations of basis elements. This means that we may write any $\beta \in \mathbb{F}_{p^n}$ in the form $\beta = \sum \beta_i f_i$, where $\{f_i\}$, $i = 1, ..., n$, is a basis, and $\beta_i \in \mathbb{F}_p$. To facilitate multiplication we define a set of numbers $\alpha_{jk}^{(m)}$ that are the expansion coefficients of the products of basis elements:

$$f_j f_k = \sum_{m=1}^{n} \alpha_{jk}^{(m)} f_m. \tag{18}$$

One may look at each $\alpha^{(m)}$ as a whole, being a symmetric $n \times n$ matrix with elements in $\mathbb{F}_p$. Having defined the α's, we may write

$$l^2 = \left(\sum_i l_i f_i\right)^2 = \sum_{jkm} l_j l_k \alpha_{jk}^{(m)} f_m \tag{19}$$

$$= \sum_m \boldsymbol{l}^{\mathrm{T}} \alpha^{(m)} \boldsymbol{l} f_m, \tag{20}$$

where $\boldsymbol{l}$ is to be thought of as a column vector whose components, $l_1, l_2, ..., l_n$, are read off the basis expansion of $l \in \mathbb{F}_{p^n}$, and $\boldsymbol{l}^{\mathrm{T}}$ is its transpose.

Now consider $\mathrm{Tr}(rl^2 + kl) = \mathrm{Tr}(rl^2) + \mathrm{Tr}(kl) = \sum_m \boldsymbol{l}^{\mathrm{T}} \alpha^{(m)} \boldsymbol{l}\, \mathrm{Tr}(rf_m) + \mathrm{Tr}(kl)$. While each trace in this last expression is difficult to calculate, we know that the trace encompasses all linear transformations from $\mathbb{F}_{p^n}$ to $\mathbb{F}_p$, and thus we may rewrite the above as

$$\mathrm{Tr}(rl^2 + kl) = \sum_m c_m \boldsymbol{l}^{\mathrm{T}} \alpha^{(m)} \boldsymbol{l} + \sum_m d_m l_m \tag{21}$$

$$= \boldsymbol{l}^{\mathrm{T}} \left(\sum_m c_m \alpha^{(m)}\right) \boldsymbol{l} + \sum_m d_m l_m, \tag{22}$$

where the d_m's (elements of $\mathbb{F}_p$) are determined by the value of k, and the c_m's (also elements of $\mathbb{F}_p$) are determined by the value of r. Here we are using the fact that every linear transformation from $\mathbb{F}_{p^n}$ to $\mathbb{F}_p$ may be written in the form $l \to \sum_m d_m l_m$ for some numbers d_m. It is convenient to condense the notation further by suppressing the sums over m. Thus we use $\mathbf{c}$ to denote the vector $(c_1, ..., c_n)$, $\boldsymbol{\alpha}$ to denote the vector of matrices $(\alpha^{(1)}, ..., \alpha^{(n)})$, and $\mathbf{c} \cdot \boldsymbol{\alpha}$ to denote $\sum_m c_m \alpha^{(m)}$. We also write $\sum_m d_m l_m = \mathbf{d}^{\mathrm{T}} \boldsymbol{l}$.

In the new notation, the nonstandard bases are given by

$$(v_{\mathbf{d}}^{(\mathbf{c})})_{\boldsymbol{l}} = \frac{1}{\sqrt{N}} e^{(2\pi i/p)[\boldsymbol{l}^{\mathrm{T}}(\mathbf{c} \cdot \boldsymbol{\alpha})\boldsymbol{l} + \mathbf{d}^{\mathrm{T}}\boldsymbol{l}]}, \tag{23}$$

where $\mathbf{c}$, $\mathbf{d}$, and $\boldsymbol{l}$ are N-component vectors with elements in $\mathbb{F}_p$. Note that the whole set of bases is determined once the n matrices $\alpha^{(1)}, ..., \alpha^{(n)}$ have been specified. It is through these matrices, defined by Eq. (18), that the structure of the field $\mathbb{F}_{p^n}$ manifests itself in the vectors v.

We know already that the above bases are mutually unbiased, but it is helpful to prove this fact directly from Eq. (23). A similar proof will be necessary when we consider the case $N = 2^n$. We begin with the inner product between vectors in different bases:

$$\begin{aligned} &|\langle v_{\mathbf{s}}^{(\mathbf{r})} | v_{\mathbf{d}}^{(\mathbf{c})} \rangle| \\ &\quad = \frac{1}{N} \left| \sum_{\boldsymbol{l}} e^{(2\pi i/p)[\boldsymbol{l}^{\mathrm{T}}\{(\mathbf{c}-\mathbf{r}) \cdot \boldsymbol{\alpha}\}\boldsymbol{l} + (\mathbf{d}-\mathbf{s})^{\mathrm{T}}\boldsymbol{l}]} \right|. \end{aligned} \tag{24}$$

If vectors from any two different bases are to be mutually unbiased, the above equation must yield the correct magnitude for all $\mathbf{c} - \mathbf{r} \neq \mathbf{0}$, and so for convenience we define the n-component vectors $\mathbf{a} = \mathbf{c} - \mathbf{r}$ and $\mathbf{b} = \mathbf{d} - \mathbf{s}$, which can take on all values with the restriction $\mathbf{a} \neq \mathbf{0}$. Now the inner product is

$$|\langle v_{\mathbf{s}}^{(\mathbf{r})} | v_{\mathbf{d}}^{(\mathbf{c})} \rangle| = \frac{1}{N} \left| \sum_{\boldsymbol{l}} e^{(2\pi i/p)[\boldsymbol{l}^{\mathrm{T}}\{\mathbf{a} \cdot \boldsymbol{\alpha}\}\boldsymbol{l} + \mathbf{b}^{\mathrm{T}}\boldsymbol{l}]} \right|. \tag{25}$$

To perform the sum, first consider the quadratic portion of the exponent,

$$\boldsymbol{l}^{\mathrm{T}}\{\mathbf{a} \cdot \boldsymbol{\alpha}\}\boldsymbol{l} = \boldsymbol{l}^{\mathrm{T}} \left(\sum_m a_m \alpha^{(m)} \right) \boldsymbol{l} = \boldsymbol{l}^{\mathrm{T}} M \boldsymbol{l}. \tag{26}$$

The action of the matrix M can be simplified if M can be diagonalized. Whenever this is the case, one obtains quadratics without cross terms in the different l_i's:

$$\boldsymbol{l}^{\mathrm{T}} D \boldsymbol{l} = \sum_m \beta_m l_i^2, \qquad D = \mathrm{diag}(\beta_1, \beta_2, ..., \beta_n). \tag{27}$$

As M is symmetric with elements in $\mathbb{F}_p$, it may always be diagonalized through a congruence transformation [12]. One can hence put $D = C^{\mathrm{T}}MC$, where C is an invertible matrix. Returning to the quadratic part of the exponent, we may write it as

$$\begin{aligned} \boldsymbol{l}^{\mathrm{T}}M\boldsymbol{l} &= \boldsymbol{l}^{\mathrm{T}}(C^{\mathrm{T}})^{-1}\, C^{\mathrm{T}}MCC^{-1}\boldsymbol{l} \\ &= \boldsymbol{l}^{\mathrm{T}}(C^{-1})^{\mathrm{T}}\, DC^{-1}\boldsymbol{l} \\ &= \mathbf{y}^{\mathrm{T}}\, D\mathbf{y}, \qquad \text{where} \quad \mathbf{y} = C^{-1}\boldsymbol{l}. \end{aligned} \tag{28}$$

We may now write the entire exponent, including the linear term, as

$$\begin{aligned} \boldsymbol{l}^{\mathrm{T}}M\boldsymbol{l} + \mathbf{d}^{\mathrm{T}}\boldsymbol{l} &= \mathbf{y}^{\mathrm{T}}\, D\mathbf{y} + \mathbf{z}^{\mathrm{T}}\mathbf{y}, \qquad \mathbf{z} = \mathbf{d}^{\mathrm{T}}C \\ &= \sum_m (\beta_m y_m^2 + z_m y_m). \end{aligned} \tag{29}$$

Finally, then, the inner product between the two vectors is

$$|\langle v_{\mathbf{s}}^{(\mathbf{r})} | v_{\mathbf{d}}^{(\mathbf{c})} \rangle| = \frac{1}{N} \left| \sum_{\mathbf{y}} e^{(2\pi i/p)[\sum_m (\beta_m y_m^2 + z_m y_m)]} \right|. \tag{30}$$

As there are no cross terms in the quadratic, we may rewrite the sum in the exponent as a product:

$$|\langle v_{\mathbf{s}}^{(\mathbf{r})} | v_{\mathbf{d}}^{(\mathbf{c})} \rangle| = \frac{1}{N} \left| \sum_{\mathbf{y}} \left(\prod_m e^{(2\pi i/p)[\beta_m y_m^2 + z_m y_m]} \right) \right|. \tag{31}$$

Further, we may sum over the y's independently, and thus the summation and product may be interchanged:

$$|\langle v_{\mathbf{s}}^{(\mathbf{r})} | v_{\mathbf{d}}^{(\mathbf{c})} \rangle| = \frac{1}{N} \prod_m \left| \left(\sum_{y_m} e^{(2\pi i/p)[\beta_m y_m^2 + z_m y_m]} \right) \right|. \tag{32}$$

And, if the β_m's are all nonzero, it follows from Eq. (12) that

$$\begin{aligned} |\langle v_{\mathbf{s}}^{(\mathbf{r})} | v_{\mathbf{d}}^{(\mathbf{c})} \rangle| &= \frac{1}{N} \prod_m \sqrt{p} = \frac{1}{N} \sqrt{p^n} \\ &= \frac{1}{\sqrt{N}}, \qquad \text{the desired result.} \end{aligned} \tag{33}$$

This last step is contingent upon the matrix $M = \sum_m a_m \alpha^{(m)}$ having no zero elements when diagonalized. In Appendix A we show this to be so for all nontrivial sets of a_m.

Unbiased Bases for Powers of 2

For the fields $\mathbb{F}_{2^n}$ the property of odd-prime fields so helpful in constructing unbiased bases [i.e., Eq. (15)] is no longer true, since for $\mathbb{F}_{2^n}$

$$\left| \sum_{j \in \mathbb{F}_{2^n}} e^{(2\pi i/2)\mathrm{Tr}[mj^2 + nj]} \right| = 0 \qquad \text{for all} \quad m, n \in \mathbb{F}_{2^n}. \tag{34}$$

We must therefore modify the above procedure, making an analog not to the simple field expression for the bases [Eq. (14)], but rather to the non-field representation described above [see Eq. (23)]. Following this idea, one can indeed obtain $N+1$ unbiased bases for $N = 2^n$, which we now write down. The standard basis is

$$(v_{\mathbf{k}}^{(\mathbf{0})})_{\boldsymbol{l}} = \delta_{\mathbf{k}\boldsymbol{l}}, \tag{35}$$

and the remaining N bases are

$$(v_{\mathbf{k}}^{(\mathbf{r})})_{\boldsymbol{l}} = \frac{1}{\sqrt{N}} i^{\boldsymbol{l}^{\mathrm{T}}\{\mathbf{r}\cdot\boldsymbol{\alpha}\}\boldsymbol{l}} (-1)^{\mathbf{k}\cdot\boldsymbol{l}}, \tag{36}$$

where $\mathbf{k}$, $\boldsymbol{l}$, and $\mathbf{r}$ are n-component vectors with elements in the set $\{0, 1\}$, and where $\boldsymbol{\alpha}$, a vector of matrices also having elements in $\{0, 1\}$, is defined as before by Eq. (18). The arithmetic in the exponent of i, however, is not to be understood as mod-2 arithmetic. That is, the sums and products indicated in that exponent are not to be done in $\mathbb{F}_2$; rather, the numbers should be treated as ordinary integers. In effect, this means that the arithmetic in that exponent is mod 4, since $i^4 = 1$. The use of the fourth roots of unity, as opposed to the square roots of unity which would be analogous to the case of odd primes, is necessary, because the square roots of unity are wholly real and do not exploit the complex nature of the space in which we are working. It is somewhat surprising that this straightforward modification of our previous formula yields a complete set of mutually unbiased bases for the present case, but it does indeed yield such a set, as we now show.

The method for showing that the above vectors are mutually unbiased is similar to our demonstration for the non-field-theoretic representation of the bases in the odd-prime case. The key is to simplify $\mathbf{a}\cdot\boldsymbol{\alpha} = \sum_m a_m \alpha^{(m)} = M$, a symmetric matrix which is the mod-4 sum of symmetric matrices $\alpha^{(m)}$ that contain only zeros and ones. One might hope that M could be diagonalized in the mod-4 arithmetic. It turns out that this is not the case in general, but one can show (Appendix B) that M can be block-diagonalized through a congruence transformation into the form

$$\mathrm{diag}(\beta_0, \beta_1, \ldots, \beta_m) \oplus q \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \oplus r \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}; \tag{37}$$

here the β_i are odd and the factors q and r indicate the number of direct sum copies to take of each matrix; thus $m + 2q + 2r = n$. The restriction on performing the block-diagonalization is that M must have an odd determinant. That this condition is met follows from the theorem of Appendix A, applied to the case $p = 2$.

As in the case of odd primes, the diagonal elements contribute separately in a multiplicative way to the final magnitude. Thus it is sufficient to show that the off-diagonal elements contribute correctly; each one encompasses two indices and thus should contribute a factor of $\sqrt{2^2} = 2$. For these indices, the exponent to which i is raised is

$$\overbrace{k_1\ k_2}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}k_1\\ k_2\end{pmatrix} = 2k_1k_2 \tag{38}$$

or

$$\overbrace{k_1\ k_2}\begin{pmatrix}2 & 1\\ 1 & 2\end{pmatrix}\begin{pmatrix}k_1\\ k_2\end{pmatrix} = 2(k_1 + k_2 + k_1k_2). \tag{39}$$

Here we have used the fact that k_1 and k_2 take only the values 0 and 1, so that $k_i^2 = k_i$. One can verify that

$$\left|\sum_{k_1,k_2=0}^{1} i^{(2k_1k_2)}\right| = \left|\sum_{k_1,k_2=0}^{1} i^{[2(k_1+k_2+k_1k_2)]}\right| = 2, \tag{40}$$

the desired result. Finally, then, we see that these bases, while less elegantly derived than those for odd primes, are nevertheless mutually unbiased. It is thus possible to find $N+1$ mutually unbiased bases for all powers of primes.

Discussion

We have seen that a complete set of mutually unbiased bases, if such a set exists for the system one is studying, provides an optimal means of determining the state of a large ensemble. Whether or not such a set exists may depend on the number N of orthogonal states available to the system. We have shown that if N is a power of a prime, then a complete set does exist. If N is not a power of a prime, then we do not know whether such a set of bases exists or not. It is clear, however, that if such bases exist at all—for example if there are seven mutually unbiased bases for a six-dimensional complex vector space—then any procedure for constructing them must be very different from the procedures we have used here, simply because our methods depend crucially on the existence of finite fields with N elements, and there is no finite field whose number of elements is not a power of a prime.

If it turns out that there are values of N for which it is not possible to construct a complete set of mutually unbiased bases—the number 6 may well be an example—then one will be faced anew with the problem of maximizing information for those cases. There will certainly be some set of $N+1$ measurements for which the information is maximized, but the measurements will not all be mutually unbiased.

We have focused our attention on the problem of state determination, but it seems likely that sets of mutually unbiased bases will be useful in other areas of physics as well. Work by one of us suggests that the Wigner-function formulation, or phase-space formulation, of quantum mechanics is intimately connected with the existence of a complete set of mutually unbiased bases [13]. The results of the present paper should therefore be relevant to the further development of this formulation for systems with a finite number of orthogonal states.

APPENDIX A

Proof that the matrix $M=\sum_m b_m\alpha^{(m)}$ is nonsingular for all nontrivial sets of b_m's: Choose a set of b_m's not all of which are zero. Let S_{mn} be an invertible matrix such that $S_{mn_0}=b_m$ for some n_0. Then the product of two basis elements in the field may be written

$$\begin{aligned} f_k f_l &= \sum_m \alpha_{kl}^{(m)} f_m = \sum_{mnp} \alpha_{kl}^{(m)} S_{mn} (S^{-1})_{np} f_p \\ &= \sum_n \beta_{kl}^{(n)} g_n, \end{aligned} \tag{A1}$$

where

$$\beta_{kl}^{(n)} = \sum_n \alpha_{kl}^{(m)} S_{mn} \qquad \text{and} \qquad g_n = \sum_p (S^{-1})_{np} f_p. \tag{A2}$$

We now do a proof by contradiction. Suppose $\sum_m b_m \alpha_{kl}^{(m)}$, which equals $\beta_{kl}^{(n_0)}$, is singular. Then it has a zero eigenvector; i.e., there exists a_l such that

$$\sum_l \beta_{kl}^{(n_0)} a_l = 0. \tag{A3}$$

Thus,

$$\sum_l f_k f_l a_l = \sum_{nl} \beta_{kl}^{(n)} a_l g_n = \sum_{n \neq n_0} \xi_n^{(k)} g_n \tag{A4}$$

for some $\xi_n^{(k)}$. The crucial fact is that the sum on the right does not involve g_{n_0}. Now define

$$h = \sum_l f_l a_l. \tag{A5}$$

Then because of Eq. (A4), we can say that *for every* k the field element $f_k h$ can be written as a linear combination of g_n's that does not include g_{n_0}. But *any* field

element x can be written as a linear combination of the f_k's, so for any x, xh can be written as a linear combination of g_n's not involving g_{n_0}. Now ask, What is g_{n_0}/h? No x will do, because xh has no g_{n_0} in it. But division must be possible because this is a field. Contradiction. Therefore, $\sum_m b_m \alpha_{kl}^{(m)}$ must not be singular. ▮

Appendix B

THEOREM. *Given an $n \times n$ symmetrix matrix M, with elements from the set $\{0, 1, 2, 3\}$ and with odd determinant, one can always find a congruence transformation in* mod-4 *arithmetic such that M is congruent to*

$$\operatorname{diag}(\alpha_0, \alpha_1, \ldots, \alpha_m) \oplus q \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \oplus r \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \tag{B1}$$

where the α_i are odd and q and r give the number of direct sum copies of each matrix. It follows that $m + 2q + 2r = n$, the rank of M.

Proof. An explicit construction [14]. The following describes a series of row and column operations which rewrite M into the desired form. The congruence transformation is thus effected by the matrix that is the product of all elementary row operations described below. The transpose of this matrix will effect the necessary column operations.

If M has odd diagonal elements. Through an interchange of rows and columns 1 and j, m_{jj}, an odd element is brought to m_{11}. Now add $-m_{l1}/m_{11}$ times row 1 to row l, and add $-m_{1l}/m_{11}$ times column 1 to column l. After doing so for all $l \geqslant 2$, we are left with

$$\begin{pmatrix} m_{11} & 0 & 0 & \cdots & 0 \\ 0 & & & & \\ 0 & & M' & & \\ \vdots & & & & \\ 0 & & & & \end{pmatrix} = m_{11} \oplus M'. \tag{B2}$$

Now consider the matrix in question to be M'. This is permissible because all row and column operations on M' will not affect rows and columns outside of M'.

If there are odd diagonal elements in M', repeat the above process for M'. Continuing this process leaves one with the direct sum of a diagonal matrix having only odd numbers on the diagonal, and a matrix having no odd diagonal elements. The latter piece may be treated as follows.

If M has no odd diagonal elements. The nonsingularity of the matrix guarantees that there is an odd m_{1l}. Interchanging rows and columns 2 and l yields one of four possibilities:

$$\begin{pmatrix} 0 & m_{12} & \cdots & \cdots \\ m_{12} & 0 & \cdots & \cdots \\ \vdots & \vdots & & \\ \vdots & \vdots & & M' \end{pmatrix}, \qquad \begin{pmatrix} 0 & m_{12} & \cdots & \cdots \\ m_{12} & 2 & \cdots & \cdots \\ \vdots & \vdots & & \\ \vdots & \vdots & & M' \end{pmatrix},$$

Case 1 Case 2 (B3)

$$\begin{pmatrix} 2 & m_{12} & \cdots & \cdots \\ m_{12} & 0 & \cdots & \cdots \\ \vdots & \vdots & & \\ \vdots & \vdots & & M' \end{pmatrix}, \qquad \begin{pmatrix} 2 & m_{12} & \cdots & \cdots \\ m_{12} & 2 & \cdots & \cdots \\ \vdots & \vdots & & \\ \vdots & \vdots & & M' \end{pmatrix}.$$

Case 3 Case 4

(a) The second and third cases can be reduced to the first: For Case 2, add row 1 to row 2 as well as column 1 to column 2. Now $m_{22} = 2m_{12} + 2$, but $2m_{12} = 2$, so $m_{22} = 2 + 2 = 0 \pmod 4$. Thus one is left with Case 1. The reduction of the third case is analogous. Having performed these reductions, one is left with either Case 1 or Case 4.

(b) Case 1: $$\begin{pmatrix} 0 & m_{12} & \cdots & \cdots \\ m_{12} & 0 & \cdots & \cdots \\ \vdots & \vdots & & \\ \vdots & \vdots & & M' \end{pmatrix}.$$

To each row $l > 2$ add $-m_{1l}/m_{12}$ times row 2 and $-m_{l2}/m_{12}$ times row 1. To each column $l > 2$ add $-m_{1l}/m_{12}$ times column 2 and $-m_{2l}/m_{12}$ times column 1. This yields, after treating all columns and rows,

$$\begin{pmatrix} 0 & m_{12} & 0 & \cdots & 0 \\ m_{12} & 0 & 0 & \cdots & 0 \\ 0 & 0 & & & \\ \vdots & \vdots & & M' & \\ 0 & 0 & & & \end{pmatrix} = \begin{pmatrix} 0 & m_{12} \\ m_{12} & 0 \end{pmatrix} \oplus M'. \tag{B4}$$

If $m_{12} = 3$, the 2×2 matrix in Eq. (B4) can be brought to the desired form by the congruence transformation

$$\begin{pmatrix} 3 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 0 & 3 \\ 3 & 0 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{B5}$$

If M' is of rank >0, begin again at step (a), replacing M with M'.

(c) Case 4: $\begin{pmatrix} 2 & m_{12} & \cdots & \cdots \\ m_{12} & 2 & \cdots & \cdots \\ \vdots & \vdots & & \\ \vdots & \vdots & & M' \end{pmatrix}$.

This cannot be reduced to Case 1 and must be treated separately. The initial manipulations are similar to Case 1: to each row $l>2$ add $-m_{l1}/m_{12}$ times row 2 and $-m_{l2}/m_{12}$ times row 1; to each column $l>2$ add $-m_{1l}/m_{12}$ times column 2 and $-m_{2l}/m_{12}$ times column 1. This procedure reduces to zero all but the first two elements in row 2 and column 2, but it may leave some unwanted 2's in row 1 and column 1. Thus the matrix is now

$$\begin{pmatrix} 2 & m_{12} & \beta_3 & \beta_4 & \cdots & \beta_n \\ m_{12} & 2 & 0 & 0 & \cdots & 0 \\ \beta_3 & 0 & & & & \\ \beta_4 & 0 & & & & \\ \vdots & \vdots & & M' & & \\ \beta_n & 0 & & & & \end{pmatrix}, \quad \text{where each } \beta_i \text{ is even.}$$

Now to each row $l>2$ add β_l/m_{12} times row 2 and to each column $l>2$ add β_l/m_{12} times column 2. This sets all but the first two elements in row 1 and column 1 to zero again without affecting row 2 and column 2. If $\beta_l=0$, m_{l2} and m_{2l} remain unchanged, and if $\beta_l=2$, $m_{l2}=2\beta_l+2(2/m_{12})=0+4(1/m_{12})=0 \pmod 4$. The argument is similar for m_{2l}.

Thus the matrix is now in the form

$$\begin{pmatrix} 2 & m_{12} & 0 & \cdots & 0 \\ m_{12} & 2 & 0 & \cdots & 0 \\ 0 & 0 & & & \\ \vdots & \vdots & & M' & \\ 0 & 0 & & & \end{pmatrix} = \begin{pmatrix} 2 & m_{12} \\ m_{12} & 2 \end{pmatrix} \oplus M'. \tag{B6}$$

If $m_{12}=3$, the 2×2 matrix can be changed to $\left(\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right)$ by means of the same congruence transformation as in step (b). If M' is of rank >0, begin again at step (a), replacing M with M'. The final result is a matrix of the form given in Eq. (B1). ∎

ACKNOWLEDGMENTS

We are grateful to David Park for a number of helpful comments on an earlier draft of the paper. Parts of the paper were inspired by unpublished work of Josephine Bellanca and Sarah Taub.

REFERENCES

1. J. SCHWINGER, *Proc. Nat. Acad. Sci. U.S.A.* **46** (1960), 570.
2. K. KRAUS, *Phys. Rev. D* **35** (1987), 3070.
3. H. MAASSEN AND J. UFFINK, *Phys. Rev. Lett.* **60** (1988), 1103.
4. I. D. IVANOVIC, *J. Phys. A* **14** (1981), 3241.
5. A general $N \times N$ complex matrix contains $2N^2$ real parameters. The requirement of Hermiticity reduces this number by a factor of 2, and the requirement of unit trace reduces it by one more, leaving $N^2 - 1$ real parameters.
6. D. V. LINDLEY, *Ann. Math. Statist.* **27** (1956), 986; W. K. WOOTTERS, "The Acquisition of Information from Quantum Measurements," p. 65, Doctoral dissertation, University of Texas at Austin, 1980.
7. B. V. GNEDENKO, "The Theory of Probability," p. 85, Chelsea, New York; 1962.
8. There is a technicality here that must be addressed. The distribution we want is the probability density of a given set of p_i's, given the observed frequencies of occurrence. The distribution we *have* (Ref. [7]) is the reverse, that is, the probability of observing a certain set of frequencies of occurrence for a given set of p_i's. These two distributions are essentially the same, as long as the a priori distribution of the p_i's is relatively slowly varying over probability space, as it is in our case.
9. A field F is a set of elements admitting two operations, addition and multiplication, with the following properties: (1) F is a commutative group under addition; (2) the nonzero elements of F form a commutative group under multiplication, where zero is the identity element for addition; (3) addition and multiplication are connected by the distributive law $a \cdot (b + c) = a \cdot b + a \cdot c$.
10. R. LIDL AND H. NIEDERREITER, "Finite Fields," p. 217, Addison-Wesley, Reading, MA, 1983.
11. For proofs and further discussion consult Lidl and Niederreiter (Ref. [10, Chaps. 1, 2]).
12. M. NEWMAN, "Integral Matrices," p. 63, Academic Press, New York, 1972.
13. W. K. WOOTTERS, *Ann. Phys. (N.Y.)* **176** (1987), 1.
14. This proof was inspired by a similar demonstration in Newman (Ref. [12, pp. 62–63]).